

Unveiling the Relationship Between Complex Networks Metrics and Word Senses

DIEGO R. AMANCIO¹, OSVALDO N. OLIVEIRA JR.¹ and LUCIANO DA F. COSTA¹

¹ *Institute of Physics of São Carlos
University of São Paulo, P. O. Box 369, Postal Code 13560-970
São Carlos, São Paulo, Brazil*

PACS 89.75.Hc – Networks and genealogical trees
PACS 02.50.Sk – Multivariate analysis
PACS 89.20.Ff – Computer science and technology

Abstract – The automatic disambiguation of word senses (i.e., the identification of which of the meanings is used in a given context for a word that has multiple meanings) is essential for such applications as machine translation and information retrieval, and represents a key step for developing the so-called Semantic Web. Humans disambiguate words in a straightforward fashion, but this does not apply to computers. In this paper we address the problem of Word Sense Disambiguation (WSD) by treating texts as complex networks, and show that word senses can be distinguished upon characterizing the local structure around ambiguous words. Our goal was not to obtain the best possible disambiguation system, but we nevertheless found that in half of the cases our approach outperforms traditional shallow methods. We show that the hierarchical connectivity and clustering of words are usually the most relevant features for WSD. The results reported here shine light on the relationship between semantic and structural parameters of complex networks. They also indicate that when combined with traditional techniques the complex network approach may be useful to enhance the discrimination of senses in large texts.

Introduction. – Many statistical methods are now used to investigate language [1] in attempts to understand empirical findings such as the Zipf's Law [2], and model syntactic and semantic relationships between words or passages [3–13]. The numerous studies encompass complex networks [14] (CN) representing texts in several applications, including summarization [5], assessment of quality of machine translators [6], authorship recognition [7], keyword extraction [8], topic identification [15] and segmentation [16]. In this paper we assess the use of complex network concepts for Word Sense Disambiguation (WSD), which is a crucial task for the Semantic Web [17] and for machine translation [18]. We analyze 10 ambiguous words with 16 topological measurements and show a strong relationship between senses and local features of complex networks. Indeed, for some of the ambiguous words the distinguishability with the CN approach is better than that obtained with the traditional analysis of neighbors.

From an analysis of feature relevance, we found that the strength of connection of neighbors in higher hierarchies and the hierarchical clustering coefficient are the most efficient metrics to discriminate word senses.

Typical Approaches to Word Sense Disambiguation. – The WSD problem has been widely studied [19] by computer scientists and researchers interested in Natural Language Processing [2] tasks. Even though humans can readily discriminate specific senses of a word, this is not the case of a computer. In fact, WSD is considered as one of the most complex problems in Artificial Intelligence [20]. The two conventional approaches to WSD are: i) the deep paradigm based on a large amount of linguist knowledge (e.g. dictionaries, thesaurus or semantic networks); and ii) the shallow paradigm which makes use of statistical techniques. The deep paradigm is in theory the best strategy as it mimics human thinking, but in practice methods requiring knowledge bases do not achieve the best perfor-

mance because there is still no database that can cover the human knowledge. Moreover, this paradigm is often impracticable because the manual creation of knowledge bases is an expensive, time consuming endeavor. In contrast, simpler methods such as those based on the analysis of contexts surrounding ambiguous words [21] have led to better performance. One of the most popular algorithms, referred to as Lesk [22], assumes that words in a given neighborhood tend to share a common topic, an assumption that is used by other algorithms [19]. Actually, the analysis of contexts based on the recurrence of nearby components is so efficient that it has even been employed to decipher encoded manuscripts [23].

Networks have also been applied to the WSD task, and some of the network-based algorithms are now close to the state-of-the-art in disambiguation. One of the earliest works date back to 1968 and uses the network structure to store knowledge in the form of a semantic memory [24]. Other examples include the application of random walks [25] in semantic networks whose nodes are linked according to semantic relations provided by WordNet [26]. With a different approach, the HyperLex algorithm [27] connect words that co-occur in a given paragraph and use the weight of edges (given by the relative frequency of occurrence of the corresponding connected nodes) to disambiguate words. Although these algorithms use the network representation in processing steps, they differ substantially from our strategy because they all consider the label of nodes while we focus on the characterization of local structural properties.

Methodology. –

Database. In the experiments, we used a set of 18 books to retrieve 10 ambiguous words (save, note, march, present, jam, ring, just, bear, rock and close), which were manually disambiguated. The only criterion in choosing these words was the quite distinct meanings of each word, which minimizes possible inaccuracies in the manual disambiguation. The list of word senses and books are given respectively in Tables S1 and S2 of the Supplementary Information (SI). The text in the books was represented as networks, as explained below.

Modeling Texts as Complex Networks. The model used to represent text is known in the literature as co-occurrence or adjacency networks [6, 7]. Basically, words are represented as nodes, which are directionally linked according to the natural reading order. In other words, if a word i appears immediately before word j in the text, then there will be the $i \rightarrow j$ edge in the network. When a given association $i \rightarrow j$ is repeated in the text, the weight of the corresponding edge is incremented. Before creating nodes and edges, stopwords (prepositions, articles and other high-frequency words with little semantic meaning) are removed (the full list of disregarded stopwords is shown in the SI). In addition, the remaining words are converted to their canonical form in order to group words with different inflections referring to a same concept.

Mathematically, the text network is defined by the matrix $W = \{w_{ij}\}$, whose element w_{ij} counts the number of times the word i appeared before the word j . When defining some of the complex networks measurements we also employed the non-oriented, non-weighted version, represented by the matrix $A = \{a_{ij}\}$ so that $a_{ij} = 1$ if i appeared at least once as a neighbor of j (regardless of the position) and $a_{ij} = 0$ otherwise. When a word repeatedly appears in the same text, it is considered as the same node in the corresponding network. But this procedure is not adopted for the ambiguous words under analysis, i.e. each occurrence is taken as a distinct node in the network, so that it is possible to characterize each occurrence of an ambiguous word to correlate its structural features with its meanings.

Characterization of Senses Through Complex Networks Features. To characterize the local structure of an ambiguous word, we used a set of 16 complex network local measurements [14]. The simplest measurement is the degree k_1 , i.e., the number of connections (without considering the weight of the edges). In terms of the adjacency matrix A , the degree is computed as

$$k_1(i) = \sum_j a_{ij}. \quad (1)$$

The weighted version of k_1 , which considers the strength [14] of the links, is given by

$$s_1(i) = \sum_j w_{ij}. \quad (2)$$

Extensions of these two measurements were considered through the analysis of further hierarchies [28] for the hierarchical expansion usually yields better network characterization [6, 28]. The expansion of a given node is made by merging the node under analysis with its neighbors in a single node, keeping the external connections of the neighbors [6, 28]. This procedure is then repeated to generate deeper expansions. This hierarchical characterization was adopted for both k_1 and s_1 , where the m -th expansion is represented as k_{m+1} and s_{m+1} . We have not made explicit use of k_{m+1} and s_{m+1} when $m = 0$ because these measurements take constant values as a consequence of considering each occurrence of an ambiguous word as a single node.

In addition to the local measurements, we quantified the connectivity of nodes to their neighbors. Analogously to the adoption of further hierarchies, the study of topological properties of neighbors also yields better network characterization [29]. Indeed, neighbors have played a key role in many algorithms, such as the PageRank [30] algorithm and its variations. In this paper, the following neighborhood-based measurements were employed: the average degree and strength of the neighbors ($\langle k_n \rangle$ and $\langle s_n \rangle$, respectively) and their standard deviations (Δk_n and Δs_n). Another structural measurement used was the clustering coefficient (C), which is proportional to the frac-

tion of triangles over the total number of connected triads. More specifically, the clustering is computed as:

$$C_1(i) = \frac{3 \sum_{k>j>i} a_{ij}a_{ik}a_{jk}}{\sum_{k>j>i} a_{ij}a_{ik} + a_{ji}a_{jk} + a_{ki}a_{kj}}. \quad (3)$$

It is known that a correlation exists between the number of semantic contexts where a word appears and its clustering coefficient [7]. Since word senses might be related to the number of contexts (because distinct senses could appear in different contexts) this measurement may be useful in the disambiguation task. Similarly to the degree and strength measurements, we expanded the hierarchies up to $m = 3$ to compute this measurement.

The local structure was also examined with shortest paths (or geodesic paths) between two nodes, which are paths whose sum of the edge weights is minimum. If d_{ij} is the shortest path between nodes v_i and v_j in the adjacency matrix A , then the average shortest l path length for v_i is:

$$l(i) = \frac{1}{n-1} \sum_{j=1}^n d_{ij}. \quad (4)$$

In networks of text, the shortest path quantifies the centrality of a word according to its distance to the most frequent words [7]. We chose to use this measurement to verify if the distance from an ambiguous word to the core-content concepts of the books can be used to distinguish senses. Shortest paths were also employed to compute the betweenness of words (B). Let η_{st}^i be the number of shortest paths between nodes v_s and v_t that pass through node v_i . If g_{st} is the number of shortest paths between nodes v_s and v_t , then B is defined as:

$$B(i) = \sum_s \sum_t \frac{\eta_{st}^i}{g_{st}}. \quad (5)$$

Even though we are aware of the correlation between B and k ($B \sim k^\eta$) [7, 31] in large networks, the possible distinct values of B taken for ambiguous words will not reflect differences in word frequency, because $k = 2$ for each occurrence of an ambiguous word. Actually, B will reflect the ability of words to connect different network regions [31, 32] or different contexts [7].

The 16 measurements employed to characterize the local structure of ambiguous words are summarized in Table 1, which contains the measurements classified according to their type (connectivity, clustering, neighborhood and paths), their notation and complexity.

To verify if the description provided by the measurements above is useful for the WSD task, we used machine learning algorithms that induce classifiers from the training set provided for each word. The quality of the results was then evaluated using the 10-fold cross validation technique [34], which was chosen because it is robust in the sense that the training set is always different from the evaluation set. Thus, it prevents that overfitted inductors

Table 1: List of measurements employed to characterize the local structure of nodes representing ambiguous words in complex networks. The last column indicates the time complexity, i.e., the time taken to compute each measurement as a function of the number of nodes n and edges e . For the clustering coefficient, the time complexity ranges between $O(n)$ and $O(n^2)$ because it depends explicitly both on $\langle k \rangle$ and $\langle k^2 \rangle$. Further details regarding measurements and time complexity can be found in Ref. [33].

Group	Notation	Complexity
Connectivity	k_2, k_3 and k_4	$O(n + e)$
	s_2, s_3 and s_4	$O(n + e)$
Clustering	C_1, C_2, C_3 and C_4	$[O(n), O(n^2)]$
Neighborhood	$\langle k_n \rangle$ and Δk_n	$O(n^2)$
	$\langle s_n \rangle$ and Δs_n	$O(n^2)$
Paths	l	$O(n^2)$
	B	$O(n^2)$

take high values of accuracy rate. 3 inductor algorithms were used: the C4.5 algorithm [35], which generates trees based on the gain provided by each feature; the Naive Bayes algorithm [35], which uses the Bayes theorem; and the k nearest neighbor algorithm [35] (kNN), which classifies an external unknown instance according to the most similar instance of the training database in a normalized space including all features. Details regarding algorithms and the cross validation technique are given in the SI.

Results and Discussion. — The 10 ambiguous words were characterized with complex networks measurements to verify if senses can be inferred from a topological analysis. Table 2 shows in the second and third columns, respectively, the accuracy rate and the corresponding p-value α_{cn} relative to a classification performed by assigning the most common (i.e. the most frequent) sense to the ambiguous word. A significant accuracy ($\alpha_{cn} < 5.0 \cdot 10^{-2}$) could be observed in 9 out of the 10 words. An example of scatter plot depicting the discrimination obtained for the word “ring” is shown in Figure 1, where each axis represents a linear combination of the 16 measurements provided by the Canonical Variable Analysis technique [36]. These results confirm the relationship between local characteristics of adjacency networks and word senses, reinforcing the suitability of complex network methods to relate structure and semantics. We believe that the ability to distinguish senses is at least partially due to the fact that co-occurrence networks probably imply syntactic factors [37] that are reflected on the semantic relations [38]. This relationship, however, is still difficult to establish because there is no consolidated interpretation for the measurements of word adjacency networks (see e.g. Ref. [7]).

Although our primary goal was not a search for the best possible disambiguation system, we compare our results with the traditional approach based on the analysis of frequency of nearby words. Classifiers were induced using attributes that represent the frequency of the 5, 20 or

Table 2: Results from characterizing ambiguous words with structural complex networks measurements. The accuracy rate and the p-value, considering as null model a classifier based on the most common sense, and the best classifier are shown for all words. The minimum set of measurements considered in each classifier is shown in the last column. Apart from the word “close”, all the classifications achieved significant accuracy rates ($\alpha_{cn} < 5.0 \cdot 10^{-2}$). Interestingly, even though 16 measurements were employed in the characterization, the top classifiers were induced with 5 measurements or less.

Word	Acc. Rate	α_{cn}	Best Ind.	Minimum Set of Measurements
save	87.64 %	$1.6 \cdot 10^{-3}$	kNN	$\langle s_n \rangle$ and l
note	84.53 %	$4.7 \cdot 10^{-2}$	kNN	C_3, s_3 and Δs_n
march	86.95 %	$1.9 \cdot 10^{-3}$	kNN	$C_3, \langle k_n \rangle, \sigma$
present	71.14 %	$2.2 \cdot 10^{-2}$	kNN	$s_4, \langle k_n \rangle, \langle k_n \rangle$ and Δs_n
jam	100.0 %	$6.0 \cdot 10^{-3}$	kNN	C_4, s_4 and Δk_n
ring	84.61 %	$2.8 \cdot 10^{-4}$	kNN	$C_4, k_4, s_4, \Delta k_n$ and l
just	51.28 %	$1.6 \cdot 10^{-2}$	kNN	C_2 and C_3
bear	61.95 %	$8.0 \cdot 10^{-3}$	kNN	C_4, k_3 and s_4
rock	79.30 %	$9.7 \cdot 10^{-9}$	C4.5	k_2, s_2, k_3, k_4 and s_4
close	72.20 %	$8.0 \cdot 10^{-2}$	kNN	$C_2, s_2, \langle s_n \rangle$ and Δs_n

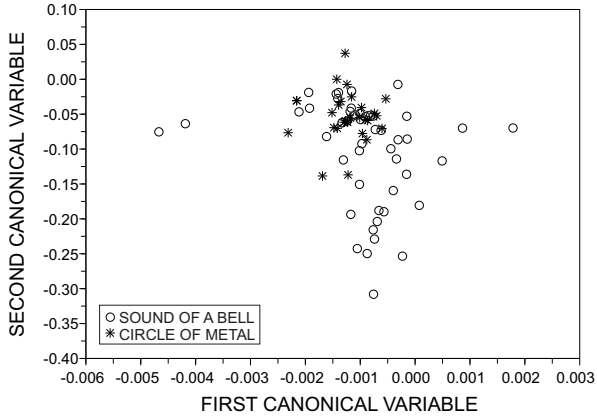


Fig. 1: Canonical Analysis Projection for the word “ring”. The senses considered were i) ring of a bell and ii) circle of metal. While the use of the sense ‘circle of metal’ is more regular as revealed by the low dispersion, the sense ‘sound of a bell’ tends to be more heterogeneous.

50 words surrounding the ambiguous word. The lowest p-values and the top classifiers are shown in Table 3. For the first 5 words the CN approach outperformed the traditional method. This means that the local structure can be even more relevant than the frequency analysis of neighbors. Obviously, we are not suggesting the CN approach to replace the approaches based on semantic information provided by neighbors, since complex network measurements are statistically reliable only when computed in large texts. Still, it could be valuable to combine both strategies in disambiguation systems. As for the methods and algorithms, the differences regarding the best classifier are worth noting. The CN approach performs better with the kNN algorithm, while the Nave Bayes algorithm is better in the traditional approach. These differences occur probably because of the distinct number of attributes

in each approach. Finally, in 6 out of the 10 words considered, the traditional analysis with 5 neighbors outperformed the classifiers with larger numbers of neighbors.

Table 3: Results from characterizing ambiguous words with the traditional approach (second column) and with the CN approach (third column). The p-value (α_{tr} for the traditional approach and α_{cn} for the CN-based method) was computed considering as null model a classifier based on the most common sense. The best classifier algorithm in the fourth column refers to the traditional approach.

Word	α_{tr}	α_{cn}	Best Ind.
save	$6.2 \cdot 10^{-1}$	$1.6 \cdot 10^{-3}$	kNN
note	$3.8 \cdot 10^{-1}$	$4.7 \cdot 10^{-2}$	kNN
march	$1.4 \cdot 10^{-2}$	$1.9 \cdot 10^{-3}$	kNN
present	$6.8 \cdot 10^{-2}$	$2.2 \cdot 10^{-2}$	Naive Bayes
jam	$1.0 \cdot 10^{-2}$	$6.0 \cdot 10^{-3}$	Naive Bayes
ring	$3.8 \cdot 10^{-9}$	$2.8 \cdot 10^{-4}$	Naive Bayes
just	$1.8 \cdot 10^{-4}$	$1.6 \cdot 10^{-2}$	Naive Bayes
bear	$3.3 \cdot 10^{-5}$	$8.0 \cdot 10^{-3}$	C4.5
rock	$< 1.0 \cdot 10^{-10}$	$9.7 \cdot 10^{-9}$	Naive Bayes
close	$< 1.0 \cdot 10^{-10}$	$8.0 \cdot 10^{-2}$	Naive Bayes

The contribution from each network metric in discriminating word senses was estimated by first finding the smallest subset of measurements generating the best classifiers (see last column of Table 2). Although we used 16 measurements to characterize nodes, the best accuracy rates were obtained with a maximum of 5 measurements. Strikingly, in some cases only two measurements were already sufficient to provide a reasonable distinction, as indicated in the scatter plot for the word “save” in Figure 2). Quantitatively, the relevance of each metric (i.e. feature) for disambiguating the words was calculated in two ways: using the Kullback-Leibler (KL) divergence and the method based on the Mann Whitney U (MWU) test [39]. While in the latter features are evaluated individually, in the former the interaction between features is considered. Thus, it is possible to identify cases where features are

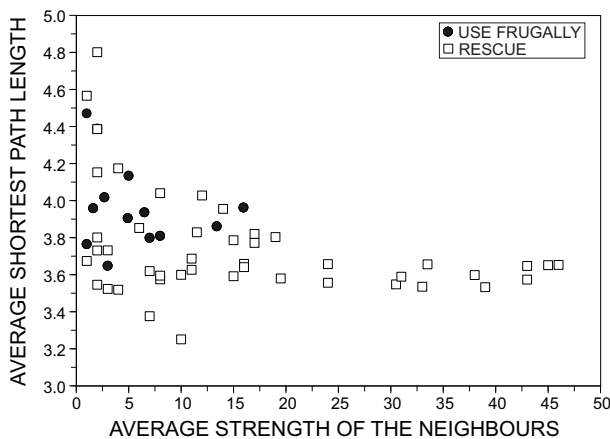


Fig. 2: Scatter plot for the word “save”. While one sense is characterized by high l and low $\langle s_n \rangle$; the other sense usually takes low values of l .

useful only when combined with others. Details regarding the KL divergence and the MWU test are given in the SI and in Ref. [7].

The rankings shown in Table 4 indicate that the relevance of a metric varies from word to word. In addition, a metric may be relevant when analyzed individually but not so if combined with other attributes because some features included in the first method may not be included in the second one. According to the KL divergence, the most frequent relevant metrics (i.e. the ones which appear among the top 3) are C_3 and l . For the MWU test, the clustering computed at high hierarchical level (C_4) is also a relevant feature along with s_3 and s_4 . Therefore, these results suggest that meanings are often correlated with the strength (or frequency) of higher-order neighbors and with the degree of interconnection of neighbors. Interestingly, l was not so relevant when combined with other features as it appeared only a few times in the MWU ranking.

Conclusion. — In this paper we have verified the suitability of the complex network model for the word sense disambiguation task in large texts. Upon characterizing the local structure of nodes representing ambiguous words, we obtained significant discrimination, which means that different senses affect the structural organization of complex networks. Strikingly, the discrimination was so effective for some words that the topological characterization outperformed traditional shallow methods. In general, the hierarchical characterization of the clustering and connectivity measurements were the most relevant features for WSD, even though the ranking of metrics varied from word to word. The analysis here may shed light on the relationship between structure of complex networks and semantics. From a practical standpoint, the methodology described might be useful in hybrid approaches to improve state-of-the-art disambiguating systems. Given an exten-

sive set of texts, it is possible to obtain networks with the local characterization of nodes representing words whose meaning is known beforehand. Then, an ambiguous word of a book could be disambiguated by assigning meanings according to the semantic (traditional approach) and topological features (CN approach) provided by the training set. In future works, we plan to use wider window sizes to connect words and additional complex networks measurements [14], such as weighted versions of the shortest path, clustering coefficient, and betweenness along with CN-based classification algorithms [40] to improve the performance of disambiguation systems in long texts. Also, we shall study the influence on the results when the proposed methodology is applied to other languages.

The authors acknowledge the financial support from CNPq (Brazil) and FAPESP (2010/00927-9).

REFERENCES

- [1] PAGE S. E., *Diversity and Complexity* (Princeton University Press) 2010.
- [2] MANNING C. D. and SCHTZE H., *Foundations of Statistical Natural Language Processing* (MIT Press) 1999, p. 24.
- [3] COSTA L. F., SPORNS O., ANTIGUEIRA L., NUNES M. G. V. and OLIVEIRA JR. O. N., *Applied Physics Letters*, **91** (2007)
- [4] FERRER I CANCHO R. and SOLE R. V., *Procs. Natl. Acad. Sci. USA*, **100** (2003), 788.
- [5] ANTIGUEIRA L., OLIVEIRA JR. O. N., COSTA L. F. and NUNES M. G. V., *Information Sciences*, **179** (2009), 584.
- [6] AMANCIO D. R., NUNES M. G. V., OLIVEIRA JR. O. N., PARDO T. A. S., ANTIGUEIRA L. and COSTA L. F., *Physica A*, **390** (2011), 131.
- [7] AMANCIO D. R., ALTMANN E. G., OLIVEIRA JR. O. N. and COSTA L. F., *New Journal of Physics* **390** (2011), 13.
- [8] MIHALCEA R., TARAU P. and FIGA E., *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, (2004), 404.
- [9] KENETT Y. N., KENETT D. Y., BEN-JACOB E. and FAUST M., *PLoS ONE* **6** 8 (2011), e23912.
- [10] SIGMAN M. and CECCHI G. A., *Proc. Natl. Acad. Sci.* **99** (2002), 1742-1747.
- [11] STEYVERS, M., TENENBAUM, J. B., *Cogn. Sci.* **29** (2005), 41-78.
- [12] ALVAREZ-LACALLE E., DOROW B., ECKMANN J.-P., MOSES E., *Proc. Natl. Acad. Sci.* **103** (2006), 7956-7961.
- [13] AMANCIO D. R., OLIVEIRA JR. O. N. and COSTA L. F., *Journal of Statistical Mechanics* **P01004** (2012).
- [14] COSTA L. F., RODRIGUES F. A., TRAVIESO G. and VILLAS BOAS P. R., *Advances in Physics*, **9** (2007).
- [15] COURSEY K., MIHALCEA R. and MOEN W., *Proceedings of the Conference on Natural Language Learning*, (2009), 210.
- [16] MALIOUTOV I. and BARZILAY R., *Proceedings of the 21st International Conference on Computational Linguistics*, (2006), 25.

Table 4: Rankings of metrics obtained with the Kullback-Leibler method and with the Mann-Whitney U test.

Method	Kullback-Leibler			Mann-Whitney U test				
	1 st	2 nd	3 rd	1 st	2 st	3 rd	4 th	5 th
bear	C_2	$\langle s_n \rangle$	$\langle k_n \rangle$	σ	s_3	k_2	k_4	Δk_n
jam	C_4	C_3	$\langle s_n \rangle$	s_3	C_4	C_2	s_4	k_4
just	C_4	C_3	—	k_3	s_4	C_4	s_3	k_4
march	—	—	—	σ	C_4	k_2	s_4	k_4
ring	l	—	—	σ	C_4	k_4	s_3	l
present	—	—	—	$\langle s_n \rangle$	C_4	$\langle k_n \rangle$	s_4	k_3
close	—	—	—	C_3	l	s_4	$\langle k_n \rangle$	k_4
note	C_3	s_4	s_2	s_3	σ	Δk_n	Δs_n	s_2
save	l	—	—	Δs_n	s_2	l	k_3	C_4
rock	l	s_2	k_2	$\langle k_n \rangle$	k_2	s_2	s_4	s_3

- [17] BERNERS-LEE T., HENDLER J. and LASSILA O., *Scientific American Magazine*, **26** (2008).
- [18] WEAVER W., *Machine Translation of Languages: Fourteen Essays* (Technology Press of MIT, Cambridge, MA, and John Wiley & Sons, New York, NY.) 1955.
- [19] NAVIGLI R., *ACM Computing Surveys*, **41** (2009) 2, 1.
- [20] MALLERY J. C., *Thinking about foreign policy: Finding an appropriate role for artificial intelligence computers.* (Ph.D. dissertation. MIT Political Science Department, Cambridge, MA.) 1988.
- [21] YARWOSKY D., *Proceedings of the ARPA Workshop on Human Language Technology*, (1993).
- [22] LESK M., *Proceedings of the 5th annual international conference on Systems documentation*, (1986), 24.
- [23] KNIGHT K., MEGYESI B. and SCHAEFER C., *Proceedings of the 4th Workshop on Building and Using Comparable Corpora*, 2011, 2.
- [24] MIHALCEA R. and RADEV D., *Graph-Based Natural Language Processing and Information Retrieval.* (Cambridge University Press) 2011.
- [25] ASZTALOS A. and TOROCZKAI Z., *Europhysics Letters*, **92**, (2010).
- [26] MIHALCEA R., TARAU P. and FIGA E., *Proceedings of the 20th International Conference on Computational Linguistics*, (2004), 1126.
- [27] VERONIS J., *Comput. Speech Lang.* **18** (2004) 3, 223.
- [28] COSTA L.F. and SILVA F. N., *Journal of Statistical Physics*, **125** (2006), 4.
- [29] COSTA L. F., RODRIGUES F. A., HILGETAG C. C. and KAISER M., *Europhysics Letters* **87**, (2009).
- [30] PAGE L., BRIN S., MOTWANI R. and WINOGRAD T., *The PageRank Citation Ranking: Bringing Order to the Web Stanford InfoLab.* (Technical Report) 1999.
- [31] BARRAT A., BARTHELEMY M. and VESPIGNANI A., *Dynamical Processes on Complex Networks.* (Cambridge University Press) 2008.
- [32] GOH K. I., KAHNG B. and KIM D., *Phys. Rev. Lett.* **87**, (2001) 278701.
- [33] NEWMAN M. E. J., *Networks: An Introduction* (Oxford University Press) 2010
- [34] KOHAVI R., *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, **2** (1995), 12.
- [35] BISHOP C. M., *Pattern Recognition and Machine Learning* (Springer) 2006
- [36] DUDA R. O., *Pattern Classification* 2000
- [37] GAMALLO P., GONZALEZ M., AGUSTINI A. LOPES G. and LIMA V. S., *Workshop on Machine Learning and Natural Language Processing for Ontology Engineering* (2002).
- [38] ROMACKER M., MARKERT K. and HAHN U., *International Joint Conferences on Artificial Intelligence* (1999), 868.
- [39] MANN H. B. and WHITNEY D. R., *Annals of Mathematical Statistics* **18** (1947) 1.
- [40] SILVA T. C. and ZHAO L., *IEEE Transactions on Neural Networks* **23** (2012), 1-16.